Plant Biotechnology Journal

Plant Biotechnology Journal (2015) **13**, pp. 200–210 doi: 10.1111/pbi.12248

Compositional differences between near-isogenic GM and conventional maize hybrids are associated with backcrossing practices in conventional breeding

Tyamagondlu V. Venkatesh[1], Kevin Cook[2], Bing Liu[2], Timothy Perez[3], Alan Willse[4], Ryan Tichich[2], Ping Feng[5] and George G. Harrigan[1,*]

[1]*Compositional Biology Center, St. Louis, MO, USA*
[2]*Trait Integration, St. Louis, MO, USA*
[3]*Statistics Technology Center, St. Louis, MO, USA*
[4]*Regulatory Affairs, Monsanto Company, St. Louis, MO, USA*
[5]*Pipeline Analytics, Monsanto Company, Ankeny, IA, USA*

Received 9 June 2014;
revised 14 July 2014;
accepted 29 July 2014.
**Correspondence* (Tel +1 314 694 8162;
fax +1 314 695 8575;
email george.g.harrigan@monsanto.com)

Keywords: Maize (*Zea mays*), marker-assisted breeding, compositional assessments, genetic modification, natural variability.

Summary

Here, we show that differences between genetically modified (GM) and non-GM comparators cannot be attributed unequivocally to the GM trait, but arise because of minor genomic differences in near-isogenic lines. Specifically, this study contrasted the effect of three GM traits (drought tolerance, MON 87460; herbicide resistance, NK603; insect protection, MON 89034) on maize grain composition relative to the effects of residual genetic variation from backcrossing. Important features of the study included (i) marker-assisted backcrossing to generate genetically similar inbred variants for each GM line, (ii) high-resolution genotyping to evaluate the genetic similarity of GM lines to the corresponding recurrent parents and (iii) introgression of the different GM traits separately into a wide range of genetically distinct conventional inbred lines. The F1 hybrids of all lines were grown concurrently at three replicated field sites in the United States during the 2012 growing season, and harvested grain was subjected to compositional analysis. Proximates (protein, starch and oil), amino acids, fatty acids, tocopherols and minerals were measured. The number of statistically significant differences ($\alpha = 0.05$), as well as magnitudes of difference, in mean levels of these components between corresponding GM variants was essentially identical to that between GM and non-GM controls. The largest sources of compositional variation were the genetic background of the different conventional inbred lines (males and females) used to generate the maize hybrids and location. The lack of any compositional effect attributable to GM suggests the development of modern agricultural biotechnology has been accompanied by a lack of any safety or nutritional concerns.

Introduction

Maize (*Zea mays* L.) is a major source of food and feed products, globally (James, 2012). Continuous improvements in traits such as herbicide tolerance, insect protection and stress tolerance are essential to the sustainable cultivation of this vital crop. Today, crop development relies extensively on the introduction of transgenic traits into plants, also referred to as genetic modification (GM). Over 70% of all maize crops are enhanced with GM traits, and this number continues to increase (James, 2012). Results from approximately 20 years of compositional studies on maize, many of which have been conducted as part of comparative safety assessments required for commercialization, have consistently shown no meaningful differences between GM crops and their conventional counterparts (Harrigan *et al.*, 2010; Herman and Price, 2013) In fact, the impact of GM trait introduction on composition is negligible relative to varietal/hybrid differences and environmental (geography, climate and agronomic practices) variability (Harrigan *et al.*, 2010; Berman *et al.*, 2011; Zhou *et al.*, 2011a,b; Harrigan and Harrison, 2012; Harrison *et al.*, 2013a,b). These results are hardly surprising and they stem from two key considerations. Firstly, domestication and breeding selection have placed some constraints on compositional variability in modern maize (Flint-Garcia *et al.*, 2009); this has led to suggestions that incorporation of more exotic alleles will be required to support continued germplasm improvements (Flint-Garcia, 2013). Secondly, both conventional breeding and the introduction of a new GM trait utilize the same type of multiple successive backcrossing steps to maximize the genetic similarity of any new line to commercially viable elite germplasm and to ensure desired agronomic characteristics (Figure 1). This backcrossing is driven primarily by breeding and agronomic considerations (Wehrhahn and Allard, 1965) but may have collateral regulatory and safety implications; as pointed out by European Food Safety Authority (EFSA) GMO Panel, 'the occurrence of unintended effects is not a phenomenon specific to genetic modification. In classical breeding extensive backcrossing, selection of favourable lines and discarding lines with unwanted properties is common practice to remove unintended effects' (EFSA, 2008).

Demonstrations of equivalence between GM and conventional counterparts have led to proposals that compositional studies are simply not required for regulatory assessments (Herman *et al.*, 2009; Herman and Price, 2013). Others have proposed that more

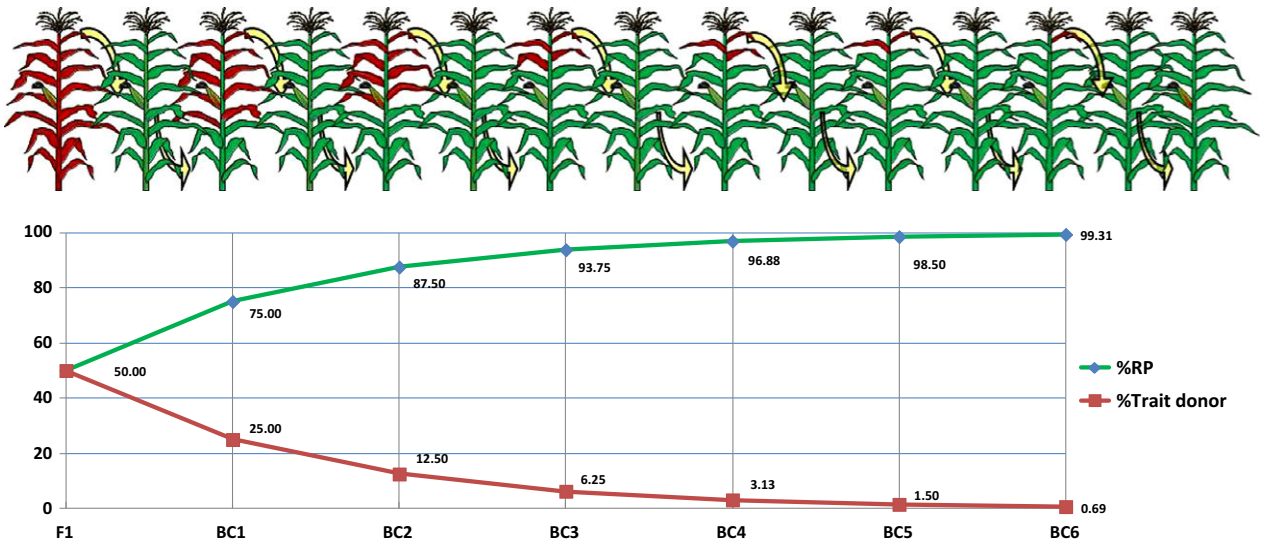

Figure 1 Overview of the successive conventional crossings involved in the GM trait integration process. Theoretically, six successive backcrosses will yield 99% genetic similarity to the desired elite germplasm. Marker-assisted backcrossing allows a reduction in the number of generations required to achieve this level of similarity. The process also generates multiple traited variants that are closely related to the elite germplasm as well as to each other.

innovative data analysis options (Harrison *et al.*, 2011, 2013a,b; Harrigan and Harrison, 2012) that acknowledge the inherent variability of crop composition would ensure that effective assessments of safety are conducted without imposing the prohibitive regulatory burdens that discriminate against smaller entrepreneurial organizations and may restrict agricultural innovation.

The purpose of this study was to provide for the first time an assessment of maize grain composition in the context of natural variability associated with conventional germplasm and with particular reference to the impact of the multiple backcrossing steps that drive the development of both conventional and GM maize products. There have been no studies on to what extent differences between GM and non-GM near-isogenic comparators are actually due to the GM trait. The study was designed specifically to distinguish quantitatively between the relative effects of a new GM trait and the residual genetic variation that distinguishes any near-isogenic lines from each other. A range of different GM traits (drought tolerance, MON 87460; herbicide resistance, NK603; insect protection, MON 89034) were included to assess the generality and robustness of study results, that is, were the observations reproducible regardless of GM trait?

Three features of the study design that are particularly informative included (i) marker-assisted backcrossing (MABC) to generate two genetically similar inbred variants for each GM line, (ii) high-resolution genotyping to evaluate the genetic similarity of GM lines to the corresponding recurrent parents during MABC and (iii) introgression of the different GM traits separately into a wide range of genetically distinct conventional inbred lines.

All F1 hybrids developed from the above breeding programme were grown concurrently at three replicated field sites in the United States during the 2012 growing season. Components analysed in the harvested grain included proximates (protein, starch and oil), amino acids, fatty acids, tocopherols and minerals, offering a comprehensive assessment of kernel composition and consistent with those used in regulatory assessments.

Results and discussion

Genetic characterization of germplasm used in maize hybrid production

The design of the experiment was founded on the use of backcrossing in developing and preserving elite germplasm. An overview of trait integration is shown schematically in Figure 1. Overall, the study included a total of four males (sometimes referred to as base inbreds) and two females (sometimes referred to as testers) to generate the GM (MON 87460, NK603 and MON 89034) hybrid sets, and selection of individuals from the backcrossing process allowed generation of near-isogenic GM variants. A summary of the hybrid and variant sets is presented in Table 1. As highlighted in Table 1, the GM trait that was to be incorporated into a hybrid was carried on either the male or female line. All traited lines were genetically fingerprinted on the Illumina (Diego, CA, USA) Infinium™ platform. The Infinium™ microarrays used for genotyping consisted of 35 000 SNPs markers. Results of the genetic similarity analysis (see Supporting Information) are presented in Table 2 and shown schematically in Figures 2 and 3. In this study, the similarity of all male and/or female inbred lines that contained a GM trait was calculated as greater than 93.7% for all comparisons to the corresponding conventional line (recurrent parent). This high degree of similarity is associated with the impact of backcrossing and is a critical feature of current conventional and GM commercial breeding practices. The genetic analysis also indicated subtle differences between the inbred variants themselves. This is the foundational concept of this study as such differences more broadly imply (i) a genetic basis for compositional differences between 'matched' variants and (ii) that differences observed between GM and non-GM comparators in many reported studies are not directly the effect of the GM trait. In other words, the results of the genetic fingerprinting allowed us to review whether compositional differences between near-isogenic GM and non-GM comparators are simply due to residual genetic variation associ-

Table 1 Overview of GM traits, hybrid sets (Female/Male) and hybrid entries (numbers 1–52)

Trait			NK603		MON 87460		MON 89034	
Female	Male	Control	A	B	A	B	A	B
A7196Z	T3653Z	1	2	3	4	5	6	7
A7196Z	T5927Z	8	9	10	11	12	13	14
A7196Z	A8389Z	15	16	17	18	19		
A7196Z	T5373Z	20			23	24	21	22
V0064Z	T3653Z	25	26	27	28	29	32	33
			30	31				
V0064Z	T5927Z	34	35	36	37	38	41	42
			39	*				
V0064Z	A8389Z	43	44	45	46	47		
V0064Z	T5373Z	48			51	52	49	50

For NK603, the GM trait was carried on the male lines except for entries 30, 31 and 39; for MON 87460, the GM trait was carried on the male line for all entries; for MON89034, the GM trait was carried on the female line except for entries 49–52. *The corresponding B variant was not available.

Table 2 Genetic similarity of traited males and female used in hybrid formation (see Materials and methods)

	% similarity
MON 87460	
T3653Z -A	98.54
T3653Z -B	97.63
T5927Z -A	94.06
T5927Z -B	94.19
A8389Z –A	96.98
A8389Z -B	97.42
T5373Z -A	99.35
T5373Z -B	97.66
NK603	
T3653Z -A	NA
T3653Z -B	97.19
V0064Z -A	98.07
V0064Z -B	98.07
T5927Z -A	98.74
T5927Z -B	95.97
A8389Z -A	97.63
A8389Z -B	97.56
MON 89034	
V0064Z -A	98.41
V0064Z -B	98.71
A7196Z -A	98.16
A7196Z -B	97.36
T5373Z -A	93.68
T5373Z -B	96.66

ated with the multiple conventional breeding steps required in the development of new GM products.

Compositional analysis

Compositional components analysed in the harvested grain included proximates (protein, starch and oil), amino acids, fatty acids, tocopherols and minerals. For all hybrids, least square mean values of each component were determined across all sites (the combined-site analysis) as well as separately for each of the three individual sites. Compositional values were consistent with those reported for maize hybrids elsewhere (Alba *et al.*, 2010). Given the large volume of data, tabulated results are presented in Tables S1–S12 and File S1. Table 3 (main text) provides a condensed overview of the data summarized by the female sets.

For the three GM traits, the following data analysis steps were taken:

1 comparisons of mean component values from the GM hybrids A and B, derived from their respective inbred variants (Table 1) to those of the conventional control hybrid derived from the respective recurrent parent as well as comparisons to each other (Table 4; Figure 4). Statistically significant differences between the mean values were declared at $\alpha = 0.05$. The purpose of this step was to compare the number of significant differences between the traited and control lines to that observed between corresponding traited variants (A and B).

2 assessment of magnitudes of differences in the comparisons of mean component values from the traited hybrids, A and B, (Table 1) with those of the conventional control hybrid derived from the respective recurrent parent (Figure 5 and File S2). This step involved determining the mean difference between each corresponding comparator at the individual sites and expressing that difference in percentages relative to the combined-site mean for the conventional control. This allowed a direct comparison of the range and distribution of component differences that could be associated with GM trait effects or with residual genetic variation.

Overall, for each trait, these two steps would elucidate any consistent trends in differences between a conventional control and the GM product when expressed in a range of genetic backgrounds. The use of a range of diverse traits would allow a robust general conclusion on the effect of genetic modification.

Finally, variance component analysis (VCA) to compare the effect of the GM trait on compositional variability relative to other experimental factors such as germplasm (the effect of different male and female lines) and location (Figure 6 and File S3) was performed across all traits.

MON 87460

MON 87460 contains a gene that encodes cold-shock protein B (CSPB) from *Bacillis subtilis*. Expression of this gene confers a yield advantage when water availability is limited (Castiglioni *et al.*, 2008). The compositional equivalence of MON 87460 to a conventional near-isogenic control has been reported (Harrigan *et al.*, 2009). In the current study, there were a total of 960 comparisons in the combined-site analysis (40 analytes × eight hybrid sets × three entries [control, hybrid A, and hybrid B] within each set). Of these, only 49 comparison (5.10%) were significantly different ($\alpha = 0.05$) and most differences were associated with comparisons between the respective hybrid variants and not between conventional and GM products. Given the paucity of observed differences, no meaningful trends were observed, that is, no analytes could be consistently associated with the GM trait, or with differences between the hybrid variants. The more meaningful interpretation of the data is that the differences between the MON 87460 and respective near-isogenic control hybrids were an unavoidable consequence of the conventional breeding steps required in developing all commercial maize, conventional or GM, that is, due to differ-

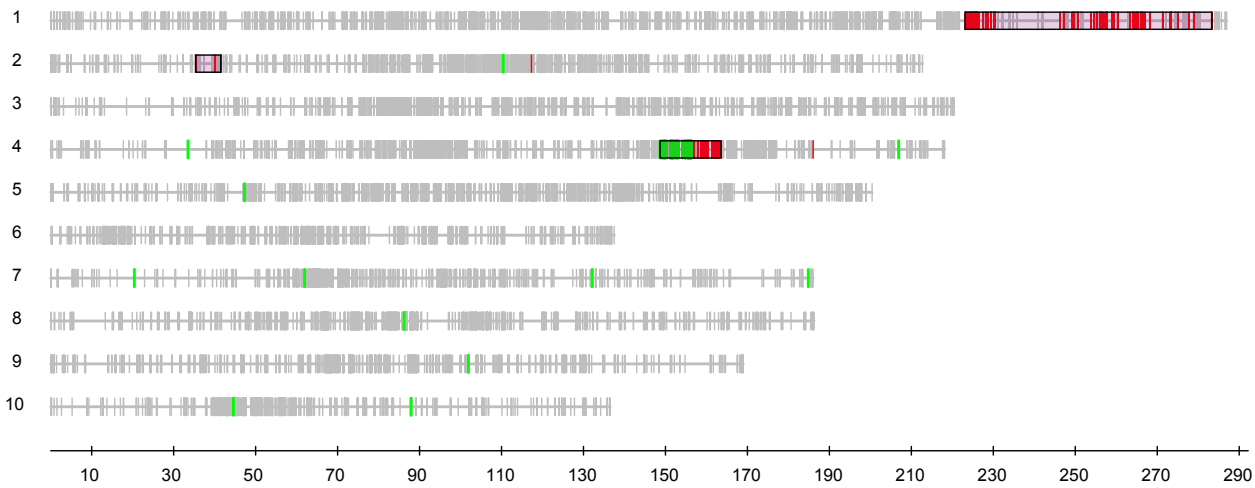

Figure 2 Comparison of the genetic profile of a MON 89034 inbred with its recurrent parent, SNP markers (vertical tics) on the ten chromosomes of the maize genetic map are coloured grey if the genotypes are the same, and either red or green if different; if recurrent parent is homozygous and the MON 89034 line is homozygous with the opposite allele, the marker is red; if recurrent parent is homozygous and the MON 89034 line is heterozygous, the marker is green. Mismatched genomic regions identified by a clustering algorithm are boxed above and account for about 4% of the genome allowing us to say the MON 89034 and recurrent parent is 96% similar. The region on chromosome 4 is near the event insertion site, corresponding to genomic segment from donor line that was selected with the event in backcross process. A large unconverted genomic region on chromosome 1 is also apparent.

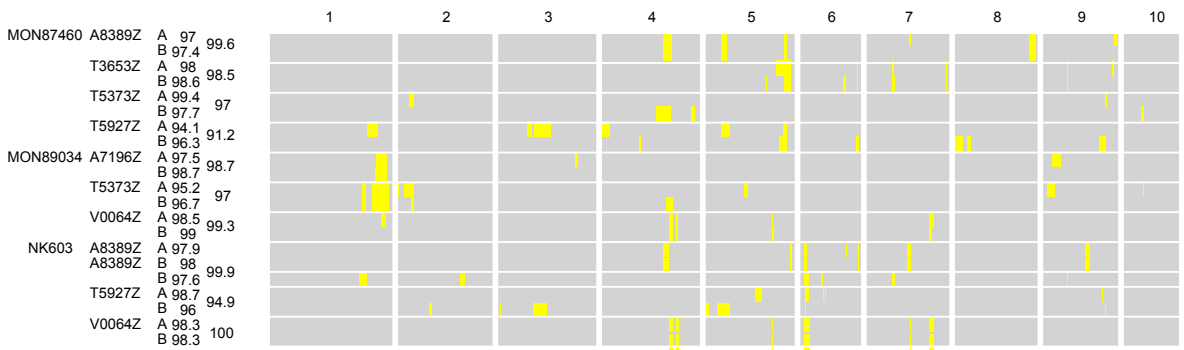

Figure 3 Comparison of all genetic profiles of GM trait-containing inbred with its recurrent parent. The number at top indicates chromosome number. They yellow segments indicated unconverted regions. The similarity of the traited inbreds to the recurrent parent as well as to each other is listed on the left of the diagram.

ences in genetic background differences rather than to the MON 87460 trait.

In assessing magnitudes of difference, it was also evident that the range and distribution of values between the GM hybrid variants or between the GM and near-isogenic conventional comparators were essentially similar. This is exemplified in Figure 5 which shows difference associated with a single component (using linoleic acid as an example, see also File S2) as well as across all components.

In summary, it can be concluded that a GM trait that acts through a regulatory protein, such as an RNA chaperone, is no more likely to impact crop composition than a trait expressed through the function of a single enzyme (such as herbicide tolerance in NK603, see below). Drought tolerance has been associated with some osmoprotectant metabolites such as sugars (e.g. trehalose) and free proline (Bartels and Sunkar, 2005), and although these do not represent a safety endpoint, their measurement could contribute to a hypothesis-driven characterization of a stress-tolerant GM product.

NK603

NK603 contains a gene that encodes the CP4 5-enolpyruvylshikimate-3-phosphate synthase (CP4 EPSPS) protein which confers tolerance to the family of Roundup® (Monsanto Company, St. Louis, MO, USA) herbicides. Earlier compositional assessments have been reported (Ridley *et al.*, 2002). In the study reported here, there were a total of 840 comparisons in the combined-site analysis (40 analytes × seven hybrid sets × three entries per set) of which only 38 (4.52%) were significantly different ($\alpha = 0.05$). As for MON 87460, no analytes could be consistently associated with the GM trait, or with differences between the variants. Assessments of magnitudes of difference were also essentially similar to that observed for the MON 87460 hybrids. This raises the question of whether nonhypothesis-driven compositional analysis is scientifically warranted for GM crop assessments. In other words, a hypothesis directly linking expected compositional changes to a GM trait would be required for meaningful interpretation otherwise

Table 3 Grain component mean values* combined by female

	Control		MON87460		MON89034		NK603	
Analyte	A7196Z	V0064Z	A7196Z	V0064Z	A7196Z	V0064Z	A7196Z	V0064Z
Proximates†								
Oil	4.09	3.81	4.13	3.82	4.09	3.75	4.27	4.06
Protein	8.36	7.71	8.52	7.86	8.41	7.72	8.52	7.84
Starch	73.95	74.12	73.91	74.16	73.94	74.28	73.68	73.69
Amino acids†								
Alanine	0.63	0.59	0.64	0.59	0.63	0.58	0.64	0.60
Arginine	0.46	0.45	0.46	0.45	0.47	0.45	0.46	0.45
Aspartate	0.57	0.54	0.58	0.54	0.58	0.54	0.57	0.55
Cysteine	0.20	0.20	0.19	0.19	0.21	0.20	0.19	0.19
Glutamate	1.73	1.61	1.75	1.62	1.75	1.59	1.74	1.64
Glycine	0.33	0.32	0.33	0.32	0.33	0.32	0.33	0.32
Histidine	0.23	0.23	0.23	0.22	0.24	0.22	0.23	0.22
Isoleucine	0.30	0.28	0.30	0.28	0.30	0.27	0.30	0.28
Leucine	1.03	0.94	1.05	0.95	1.04	0.93	1.04	0.96
Lysine	0.25	0.25	0.25	0.25	0.26	0.25	0.25	0.25
Methionine	0.19	0.17	0.18	0.17	0.19	0.18	0.18	0.17
Phenylalanine	0.35	0.32	0.35	0.32	0.35	0.32	0.35	0.33
Proline	0.80	0.77	0.79	0.77	0.80	0.75	0.78	0.77
Serine	0.43	0.41	0.44	0.41	0.44	0.41	0.43	0.41
Threonine	0.31	0.29	0.31	0.29	0.31	0.29	0.31	0.29
Tryosine	0.35	0.33	0.36	0.33	0.37	0.33	0.35	0.33
Tryptophan	0.061	0.062	0.062	0.063	0.063	0.061	0.062	0.062
Valine	0.42	0.40	0.42	0.40	0.43	0.40	0.42	0.40
Fatty acids‡								
Palmitic	11.48	11.08	11.26	11.19	11.91	11.29	11.38	10.96
Stearic	1.74	1.84	1.76	1.85	1.71	1.86	1.74	1.82
Oleic	28.83	27.67	28.52	27.78	28.04	27.04	29.66	28.48
Linoleic	55.23	56.61	55.71	56.45	55.63	57.02	54.53	56.04
Linolenic	1.49	1.50	1.49	1.46	1.47	1.52	1.48	1.46
Eicosenoic	0.42	0.46	0.43	0.46	0.45	0.45	0.43	0.43
Eicosadienoic	0.34	0.35	0.33	0.34	0.33	0.34	0.33	0.34
Behenic	0.15	0.14	0.15	0.14	0.16	0.15	0.12	0.14
Tocopherols§								
α-Tocopherol	14.10	11.11	13.74	11.81	12.79	11.74	12.10	10.82
δ-Tocopherol	1.33	0.74	1.34	0.79	1.55	0.98	1.52	0.96
γ-Tocopherol	42.50	31.07	42.01	31.19	46.52	35.96	46.15	35.27
Minerals¶								
Aluminum	41.68	43.43	41.91	43.51	44.87	42.48	43.26	41.52
Calcium	0.0035	0.0040	0.0034	0.0034	0.0034	0.0038	0.0035	0.0034
Iron	15.39	14.90	15.58	14.70	15.50	14.04	15.61	14.30
Magnesium	0.081	0.087	0.081	0.084	0.082	0.085	0.081	0.086
Manganese	4.45	3.60	4.46	3.49	4.74	3.70	4.41	3.46
Phosphorus	0.22	0.22	0.22	0.21	0.22	0.21	0.22	0.22
Potassium	0.27	0.27	0.27	0.27	0.27	0.27	0.27	0.27
Zinc	15.20	15.25	14.58	14.36	15.51	15.02	16.28	14.23

*Least square means determined across all three individual sites and males. Means and standard errors for individual hybrids are shown in Tables S1–S12.

†Expressed as % dwt, starch, protein determined by NIR,

‡Fatty acids expressed as % total FA,

§Expressed as mg/kg dwt.

¶Aluminum, iron, manganese, zinc expressed as mg/kg dwt, calcium, magnesium, phosphorus, potassium expressed as % dwt.

observed differences between two near-isogenic comparators can most realistically be assumed to be due to residual genetic variation. In this context, it is noteworthy that EFSA (2008) states that animal studies are warranted only when compositional differences are seen for the GM effect yet NK603 has been associated with recent controversy related to highly disputed results reported in a formerly retracted (now republished) rat feeding study (Casassus, 2014).

Table 4 Number of statistically significant differences (a = 0.05) from the combined-site analysis (three field sites)

Trait	Hybrid A vs B	Control vs A	Control vs B	Total
MON 87460	19*	13*	17*	49 (5.0%)†
NK603	11‡	18‡	9‡	38 (4.52%)§
MON89034	12¶	18¶	12¶	42 (5.83%)**
Total	42/840 (5.00%)	49/840 (5.83%)	38/840 (4.52%)	129/2520 (5.11%)

*From a total of 320 comparisons.
†From a total of 960 comparisons.
‡From a total of 280 comparisons.
§From a total of 840 comparisons.
¶From a total of 240 comparisons.
**From a total of 720 comparisons.

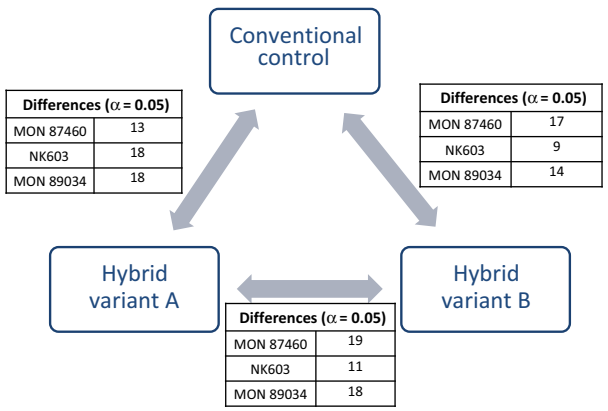

Figure 4 A three-way comparison of the hybrid variants and the corresponding conventional control. Differences in comparisons represented by the diagonal arrows can be attributed to residual genetic variation; differences in comparisons represented by the horizontal arrow can be attributed to residual genetic variation. Components identified as being statistically significantly different are presented in the Supporting Tables.

MON 89034

MON 89034 expresses two Cry insecticidal proteins (Cry1A.105 and Cry2ab2) from *Bacillis thuringiensis* that provide protection against lepidopteran insect pests. The compositional equivalence of MON 89034 to a conventional near-isogenic control has been reported (Drury *et al.*, 2008). In this study, only 42 (5.83%) of 720 comparisons in the combined-site analysis (40 analytes × six hybrid sets × three entries per set) were significantly different ($\alpha = 0.05$). Overall, as for the other traits, no analytes could be consistently associated with the GM trait, or with differences between the variants. In other words, results obtained in this study for the three distinct traits, drought tolerance, herbicide resistance, and insect protection, were near-identical.

With respect to MON 89034, compositional studies on other Cry protein-containing products such as MON 810 have confirmed that differences between GM and non-GM comparators are of similar number and magnitude to that consistent with the residual genetic variation associated with the backcrossing steps used in the generation of conventional and GM products (Zhou *et al.*, 2011b). It can be extrapolated that meaningful compositional changes are unlikely to be induced in any Cry protein-containing products.

Variance component analysis

As mentioned earlier, magnitudes of differences in the comparisons of mean component values from the traited hybrids, A and B, with those of the conventional control hybrid derived from the respective recurrent parent were evaluated. This step involved determining the mean difference between each corresponding comparator at the individual sites and expressing that difference in percentages relative to the combined-site mean for the conventional control. At the same time, the model was used to estimate the means difference between the same conventional control hybrid grown in different sites and provided information on the impact of environment on composition (e.g. Figure 5). To more effectively compare the relative contributions of environment, germplasm, trait, and residual genetic variation, a variance component analysis (VCA) was performed. This analysis confirmed a lack of 'trait' effect; in this context, 'trait' effect could more reasonably be referred to as a 'variant' effect because it refers to differences between A and B variants from the control and therefore includes a contribution from both the GM trait and background genetic effects. The result of VCA combined across all components and all three GM traits is presented in Figure 6. Data presented in Supporting Information show that variation in levels of proximates, amino acids, fatty acids, and tocopherols were generally dominated by germplasm (male–female combination), location, and residual effects. Variation in minerals was dominated, in general, by location and residual effects. The trait effect (or variant effect) was essentially zero for all components. The impact of germplasm and environment on compositional variation is now well established in studies on GM crops (Harrigan *et al.*, 2010; Berman *et al.*, 2011; Zhou *et al.*, 2011a,b; Harrigan and Harrison, 2012; Harrison *et al.*, 2013a,b), and it is increasingly evident that GM is a negligible contributor to that variation.

Conclusion

There are numerous rigorous selection and quality control steps that ensure a lack of unintended effects in newly developed GM crops (Privalle *et al.*, 2012). Moreover, the development of new GM maize hybrids is, at its core, a conventional breeding venture involving extensive backcrossing steps to ensure a high degree of genetic similarity to elite, high performing commercial germplasm. By generating GM variants that are as near-isogenic as practically possible to the respective conventional control, as well as to each other, our study design allowed us to distinguish potential GM effects on composition from residual genetic variation associated with backcrossing. Results showed that differences that would be observed in comparisons between any near-isogenic comparators, conventional or GM, greatly exceed that of GM effects and provide a strong underpinning as to why the impact of GM technology on composition has been consistently found to be negligible. This observation extended not only to physiologically simple single gene traits such as herbicide tolerance or insect protection but to traits that enhance plant physiology to better enable yield or stress tolerance, traits that have also been sought through conventional breeding selection.

It is also evident that the term *genetic modification* is a misnomer when referring to crops developed through modern biotechnology (Herring, 2008). GM traits can be introgressed into a wide range of conventional germplasm without impacting

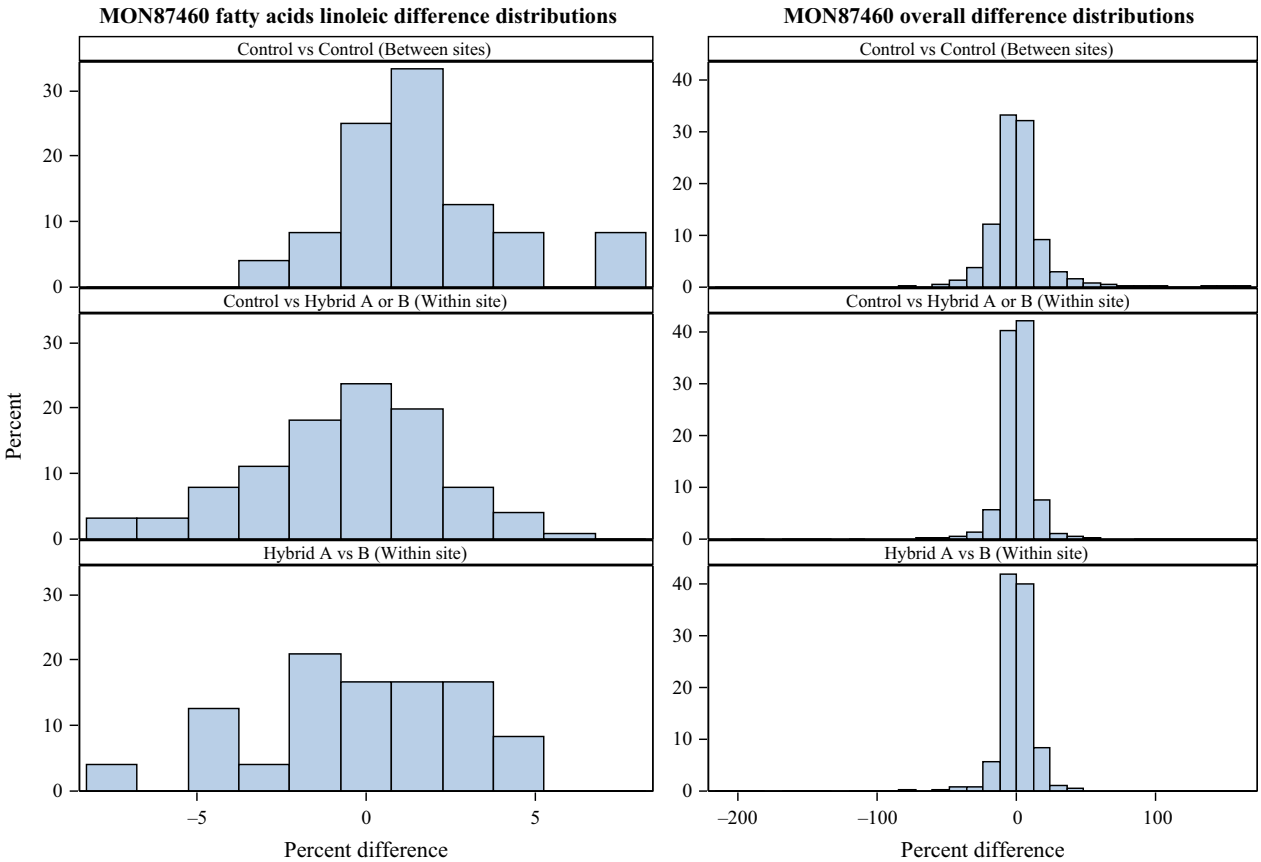

Figure 5 The plots on the left show (bottom caption) magnitudes of difference observed for all corresponding A versus B comparisons expressed as percentage difference, (middle caption) magnitudes of differences observed of A and B hybrids versus control and (top caption) differences between control hybrids grown at different sites. The plots on the right are similar plots combined across all components.

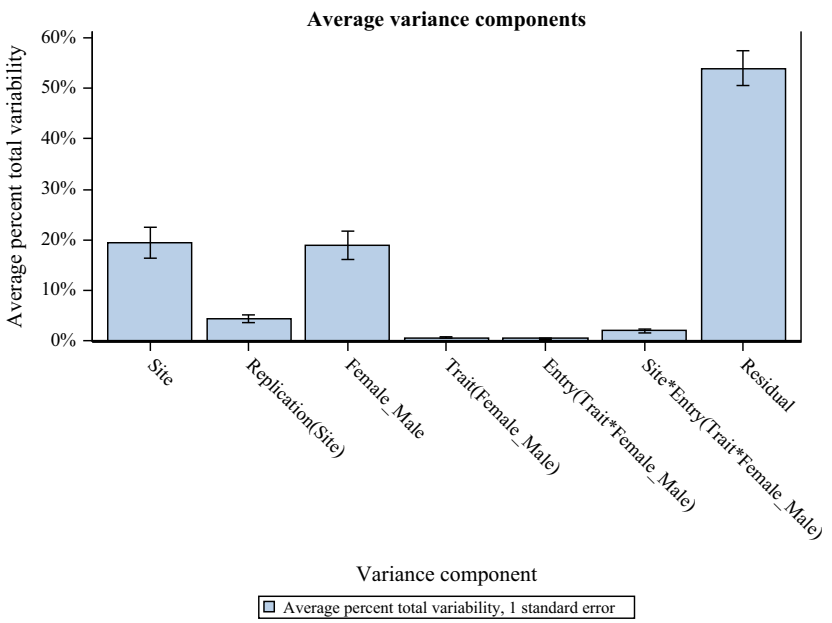

Figure 6 Variance component analysis averaged across all traits and all hybrids. These results highlight the lack of any trait effect; the term Trait (Female_Male) represents variation due to trait (Control, MON87460, MON89034 and NK603) within a female and male combination but does include a contribution from residual genetic variation. The term Entry (Trait × Female_Male) represents variation due to differences between the A and B variants (i.e. residual genetic variation).

extant levels of genetic quality and diversity; the binary classification of GM and non-GMO crops is an artificial dichotomy from the perspective of plant composition and genetics.

Our results have implications for current practices and principles of current safety assessments. The results of previous studies which have assigned differences between near-isogenic conventional and GM comparators to the GM trait may have to be reconsidered. As has been pointed (Herman and Price, 2013), the lack of unintended compositional consequences observed after decades of studies on a range of GM crops support the

safety of the GM process. Compositional studies can therefore only have merit if there is a specific hypothesis on effects that can be directly associated with insertion of a given GM trait. In other words, issues related to safety and nutrition can be more effectively addressed through targeted hypothesis-driven evaluations than by large-scale prescriptive studies that fail to distinguish residual genetic variation between near-isogenic comparators. This is especially true when considering hypothesis-free nontargeted profiling approaches that can neither associate observed differences with the GM trait nor a clear safety or nutritional endpoint. The lack of any compositional effect comprehensively established here for a range of diverse GM traits suggests that citing the so-called precautionary principle to limit the development of modern agricultural biotechnology is misplaced and that the expense and resources associated with precautionary approaches are preventing beneficial traits being developed by smaller organizations that lack the resources to generate required regulatory data.

Materials and methods

Maize samples, nursery and field production

In summary, there were a total of four males (Manufacturer Codes: T3653Z, T5373Z T5927Z and A8389Z) and two females (Manufacturer Codes: A7196Z and V0064Z) (Table 1). These males and females provided the foundation for a range of different traited and conventional control hybrids in this study. The three GM traits chosen for the study included stress tolerance, (MON 87460) (Castiglioni *et al.*, 2008), herbicide tolerance, (NK603) (Ridley *et al.*, 2002), and insect protection, (MON 89034) (Drury *et al.*, 2008). For each GM trait, sets of hybrid variant (A and B) were generated from respective base inbred (male) or tester (female) variants (A and B). Multiple hybrids were generated for each GM trait and each variant, in three to four different genetic backgrounds depending on the GM trait. Control hybrids used for comparison were generated in same genetic background but without the GM trait. A total of 51 hybrids were generated for analysis (Table 1).

Production of transgenic conversions

The conversions used in this experiment were produced by standard backcross breeding supplemented with molecular markers (Eathington *et al.*, 2007). Tissue samples for each individual seedling growing in the conversion nursery were shipped to the Monsanto marker laboratory in Ankeny, Iowa, where they were assayed with PCR using 100 polymorphic markers. Individuals most closely related to the recurrent parent for each conversion were pollinated and advanced to the next generation. This was continued for three to five generations depending on the inbred. After the last backcross, plants were self-pollinated to produce homozygous lines and further increased to produce F4 seed bulks to be used in hybrid production.

The A and B versions were developed by selfing two different plants in the last backcross generation BC3 generation and continuing two separate lineages from that point.

Genetic fingerprint analysis

A bulked seed sample from each of the finished conversions was shipped to the genotyping laboratory at Monsanto St Louis for fingerprinting. The conversions were fingerprinted using the Illumina Infinium™ platform. The Infinium™ microarrays used for genotyping consisted of 35 000 SNPs markers. The SNPs are proprietary Monsanto genetic markers, mostly discovered from the sequencing of two public lines: MO17 and B73 and two elite Monsanto inbreds. By comparing or overlaying genotypes of GM conversions with parental line on the genetic map, we can identify genomic regions where they differ. For example, Figure 2 compares a MON 89034 conversion with its recurrent parent, colouring SNP markers (vertical tics) on genetic map grey if their genotypes are the same, and either red or green if their genotypes are different; if recurrent parent is homozygous and conversion is homozygous with the opposite allele, marker is coloured red; if recurrent parent is homozygous and conversion is heterozygous, marker is coloured green. As expected, and as exemplified in Figure 2, mismatched SNPs tend to cluster, with the clusters defining continuous genomic segments where the lines differ. In Figure 2, the region on chromosome 4 is near the event insertion site, corresponding to genomic segment from donor line that was selected with the GM transgene in backcross process. A large unconverted genomic region on chromosome 1 is also apparent.

To delimit contiguous regions on the genetic map where the near isolines differ, we used a clustering algorithm that identified groupings of improbably neighbouring mismatched markers, accounting for the probability that some SNPs within a region of difference will match (i.e. identical by state but not identical by descent), variability in marker map positions, as well as fingerprint error rate. Delimiting genomic regions is complicated by the large number of matching SNPs within a region. In a straightforward backcross conversion with a single donor and a single recurrent parent, this complication can be avoided by first identifying markers whose genotypes differ between the donor and recurrent parent line; these are the so-called informative markers. When the clustering algorithm is applied using informative markers only, the genomic segments can be more precisely defined. In related work, Frisch and Melchinger (2006) used low-density genetic fingerprints of donor, recurrent parent and offspring to attribute genomic regions in offspring to donor or recurrent parent. Their application required knowledge of informative markers, and extensive interpolation due to low marker density.

Hybrid production

The hybrid seed was produced at the Monsanto research farm in Kihei, Hawaii. The parents (male and female inbreds) for each hybrid were planted in three rows called 'triplets' where the male parent was planted in the centre row flanked by two rows of the same female parent. This configuration was used for each hybrid in the study. To produce the seed, pollen from the male parent in the centre row was collected in standard paper pollinating bags and transferred to the silks of the female parents in the adjoining rows by holding the pollinating bag over the silks and shaking the bag, so the cloud of pollen would settle on all of the exposed silks. To prevent other pollen from contaminating the sample, a 'shoot bag' was kept over the ear shoot until pollination and the pollinating bag was left over the ear after pollination until harvest. The ears were harvested after maturity, dried to approximately 12% moisture in the seed and shipped to the Monsanto research facility in Huxley Iowa for processing and packaging for planting.

Field trials

The hybrids were planted in the field in 2012 growing season in a randomized complete block design at three different locations in United States (Boone County, Iowa and Sangamon and McClean Counties, Illinois). At each location, the plots comprised four rows (7 m long and 0.7 m between rows). Each hybrid material was planted in three replications at each location. Standard agro-

nomic practices for each geographic region were followed. At harvest, six ears were harvested by hand and were dried down to 14% moisture before shelling the seed form the cobs. The seed for the six ears was bulked and shipped to St Louis for compositional analysis.

Analytical methods

Grain samples were analysed for protein, oil, starch, amino acids, fatty acids, minerals and tocopherols. Moisture levels were measured for the re-expression of fresh weight values on a dry weight basis. Values for protein, oil and starch were determined using NIRS (near-infrared spectroscopy) following ISO (International Standard Organization) certified methods. Approximately 250-g sample of whole-kernel was used in a Foss Infratec 1221 near-infrared transmittance (NIRT) instrument. The NIRT was calibrated using reference wet chemistry methods. %RSD (relative standard deviation) calculated from control samples for protein, oil and starch were 1.49%, 3.02% and 0.71%, respectively. Each test sample was measured in duplicate to verify repeatability and average of the two repeats was reported. Where replicate analyses were more than 3 × standard deviation, a third replicate was run. When there were more than two replicate values, a Q-Test was performed to determine whether there is an outlier. If Q (observed) > Q (tabulated), the outlier is discarded and average of the only the two remaining results were reported. All NIRT results were reported on a dry basis percentage (percentage of nonwater material).

Methods for determination of amino acids were as described previously in Harrigan *et al.* (2007). Fatty acids were determined using AOCS (American Oil Chemists' Society) methods as described in Harrigan *et al.* (2007). Minerals (calcium, copper, iron, magnesium, manganese, phosphorus, potassium, sodium and zinc) levels were estimated using inductively coupled plasma emission (ICP) mass spectrometry-based AOAC methods as described in Drury *et al.* (2008). Tocopherol analysis was based on a reversed-phase HPLC method using fluoresence detection with excitation at 290 nm and emission at 336 nm (Hogarty *et al.*, 1989). Tocopherols were extracted from ground lyophilized seed with 0.1% pyrogallol in ethanol. The reversed-phase HPLC system comprised a Keystone Aquasil C_{18} column at 40 °C, and methanol as mobile phase. Flow rate was 1 mL/min.

Statistical analysis

Statistical analyses were performed using SAS Software (SAS Institute, Cary, NC, USA) Release 9.4. All compositional components were statistically analysed using a mixed model analysis of variance. The three replicated sites were statistically assessed individually (individual site analysis) and as a combination of the all three sites (combined-site analysis).

Combined-site analysis was performed using the following model:

$$Y_{ijklm} = U + S_i + R(S)_{ij} + H_k + T(H)_{kl} + E(\mathrm{TH})_{klm} + \mathrm{SE(TH)}_{iklm} + e_{ijklm}$$

where Y_{ijklm} is the unique individual observation, U is the overall mean, S_i is the random site effect, $R(S)_{ij}$ is the random replicate within site effect, H_k is the hybrid or male × female effect, $T(H)_{kl}$ is the trait within hybrid effect, $E(\mathrm{TH})_{klm}$ is the entry within trait and hybrid combination effect, $\mathrm{SE(TH)}_{iklm}$ is the random site by entry within trait and hybrid combination interaction effect, and e_{ijklm} is the residual error.

Individual site analysis was performed using the following model:

$$Y_{ijkl} = U + R_i + H_j + T(H)_{jk} + E(\mathrm{TH})_{jkl} + e_{ijkl}$$

where Y_{ijkl} is the unique individual observation, U is the overall mean, R_i is the random replicate effect, H_j is the hybrid or male × female effect, $T(H)_{jk}$ is the trait within hybrid effect, $E(\mathrm{TH})_{jkl}$ is the entry within trait and hybrid combination effect, and e_{ijkl} is the residual error.

The SAS procedure PROC MIXED was employed to run the analysis. A residual is the difference between the observed value and its predicted value from a statistical model. A studentized residual is scaled so that the residual values tend to have a standard normal distribution when outliers are absent. Thus, most values are expected to be between ±3. Extreme data points that are also outside of the ±6 studentized residual range are considered for exclusion, as outliers, from the final analyses. A total of 18 observations of 18 360 had studentized residuals outside of the ±6 range (one cysteine, one glycine, three eicosadienoic, five iron, one magnesium, one manganese, one zinc, three γ-tocopherol and two δ-tocopherol). These observations were identified as outliers and removed from analysis.

For each compositional component, the mean of the control hybrid was compared to the mean of the respective A and B inbred variants of each trait. Likewise, for each compositional component and trait, the mean of the respective A and B inbred variants was compared to each other. Statistically significant differences between the mean values were declared at $\alpha = 0.05$.

Assessment of magnitudes of difference was performed using the combined-site ANOVA model with the effects of site and the interaction between site and entry within trait and hybrid combination as fixed effects. The model was used to estimate the within site means difference between each of the traited hybrids, A and B, (Table 1) and the conventional control hybrid derived from the respective recurrent parent for each compositional component. At the same time, the model was used to estimate the means difference between the same conventional control hybrids grown in different sites. All differences were expressed as percentages relative to the combined-site mean for the conventional control. This step therefore allowed a direct comparison of the range and distribution of component that could be associated with GM trait effects or with residual genetic variation.

Variance components analysis (VCA) was also conducted to estimate the relative contribution of the experimental factors to the total variance in the study. In this application, all effects from the combined-site ANOVA model were set as random effects. The SAS procedure PROC MIXED was employed to run the analysis. The output table of covariance parameter estimates from SAS PROC MIXED procedure gives estimates of the variance component parameters for each of model components. The variance component parameters of each model component were divided by the total variance to obtain the variance proportions for each component.

Acknowledgements

We are very grateful for the agronomic support provided by all our colleagues in the breeding organizations at Monsanto. The excellent logistical support provided by Sami-Abdul Samad, Kimberlee Shaeffer in the Sample Management Team, and the

statistical programming support provided by Susan Riordan were also critical to the success of this study. We are indebted to Nordine Cheikh for his support and encouragement and grateful to Kevin Glenn, John Vicini and Angela Hendrickson Culler for helpful comments of the original drafts.

Conflict of interest

There is no direct conflict of interest, but the authors are employed by the agricultural biotechnology industry.

References

Alba, R., Phillips, A., Mackie, S., Gillikin, N., Maxwell, C., Brune, P., Ridley, W., Fitzpatrick, J., Levine, M. and Harris, S. (2010) Improvements to the international life sciences institute crop composition database. *J. Food Compost. Anal.* **23**, 741–748.

Bartels, D. and Sunkar, R. (2005) Drought and salt tolerance in plants. *Crit. Rev. Plant Sci.* **24**, 23–58.

Berman, K.H., Harrigan, G.G., Nemeth, M.A., Oliveira, W.S., Berger, G.U. and Tagliaferro, F.S. (2011) Compositional equivalence of insect-protected glyphosate-tolerant soybean MON 87701 × MON 89788 to conventional soybean extends across different world regions and multiple growing seasons. *J. Agric. Food Chem.* **59**, 11643–11651.

Casassus, B. (2014) Paper claiming GM link with tumours republished. Change of journal does not convince critics that rat diseases were caused by genetically modified maize. *Nature*, **511**, 10. doi:10.1038/nature.2014.15463.

Castiglioni, P., Warner, D., Bensen, R.J., Anstrom, D.C., Harrison, J., Stoecker, M., Abad, M., Kumar, G., Salvador, S., D'Ordine, R., Navarro, S., Back, S., Fernandes, M., Targolli, J., Dasgupta, S., Bonin, C., Luethy, M.H. and Heard, J.E. (2008) Bacterial RNA chaperones confer abiotic stress tolerance in plants and improved grain yield in maize under water-limited conditions. *Plant Physiol.* **147**, 446–455.

Drury, S.M., Reynolds, T.L., Ridley, W.P., Bogdanova, N., Riordan, S., Nemeth, M.A., Sorbet, R., Trujillo, W.A. and Breeze, M.L. (2008) Composition of forage and grain from second-generation insect-protected corn MON 89034 is equivalent to that of conventional corn (*Zea mays* L.). *J. Agric. Food Chem.* **56**, 4623–4630.

Eathington, S.R., Crosbie, T.M., Edwards, M.D., Reiter, R.S. and Bull, J.K. (2007) Molecular markers in a commercial breeding program. *Crop Sci.* **47**, S154–S163.

EFSA (2008) Safety and nutritional assessment of GM plants and derived food and feed: the role of animal feeding trials. *Food Chem. Toxicol.* **46** (Suppl. 1), S2–S70.

Flint-Garcia, S.A. (2013) Genetics and consequences of crop domestication. *J. Agric. Food Chem.* **61**, 8267–8276.

Flint-Garcia, S.A., Bodnar, A.L. and Scott, M.P. (2009) Wide variability in kernel composition, seed characteristics, and zein profiles among diverse maize inbreds, landraces, and teosinte. *Theor. Appl. Genet.* **119**, 1129–1142.

Frisch, M. and Melchinger, A.E. (2006) Marker-based prediction of the parental genome contribution to inbred lines derived from biparental crosses. *Genetics*, **174**, 795–803.

Harrigan, G.G. and Harrison, J.M. (2012) Assessing compositional variability through graphical analysis and Bayesian statistical approaches: case studies on transgenic crops. *Biotechnol. Genet. Eng. Rev.* **28**, 15–32.

Harrigan, G.G., Stork, L.G., Riordan, S.G., Reynolds, T.L., Ridley, W.P., Masucci, J.D., MacIsaac, S., Halls, S.C., Orth, R., Smith, R.G., Wen, L., Brown, W.E., Welsch, M., Riley, R., McFarland, D., Pandravada, A. and Glenn, K.C. (2007) Impact of genetics and environment on nutritional and metabolite components of maize grain. *J. Agric. Food Chem.* **55**, 6177–6185.

Harrigan, G.G., Ridley, W.P., Miller, K.D., Sorbet, S., Riordan, S.G., Nemeth, M.A., William Reeves, W. and Pester, T.A. (2009) The forage and grain of MON 87460, a drought-tolerant corn hybrid, are compositionally equivalent to that of conventional corn. *J. Agric. Food Chem.* **57**, 9754–9763.

Harrigan, G.G., Lundry, D., Drury, S., Ridley, W.P., Riordan, S.G. and Nemeth, M.A. (2010) Natural variation in crop composition and the impact of transgenesis. *Nat. Biotechnol.* **28**, 402–404.

Harrison, J.M., Breeze, M.L. and Harrigan, G.G. (2011) Introduction to Bayesian statistical approaches to compositional analyses of transgenic crops 1. Model validation and setting the stage. *Regul. Toxicol. Pharmacol.* **60**, 381–388.

Harrison, J.M., Howard, D., Malven, M., Halls, S.C., Culler, A.H., Harrigan, G.G. and Wolfinger, R.D. (2013a) Principal variance component analysis of crop composition data: a case study on herbicide-tolerant cotton. *J. Agric. Food Chem.* **61**, 6412–6422.

Harrison, J.M., Breeze, M.L., Berman, K.H. and Harrigan, G.G. (2013b) Bayesian statistical approaches to compositional analyses of transgenic crops 2. Application and validation of informative prior distributions. *Regul. Toxicol. Pharmacol.* **65**, 251–258.

Herman, R.A. and Price, W.D. (2013) Unintended compositional changes in genetically modified (GM) crops: 20 years of research. *J. Agric. Food Chem.* **61**, 11695–11701.

Herman, R.A., Chassy, B.M. and Parrott, W. (2009) Compositional assessment of transgenic crops: an idea whose time has passed. *Trends Biotechnol.* **27**, 555–557.

Herring, R.J. (2008) Opposition to transgenic technologies: ideology, interests and collective action frames. *Nat. Rev. Genet.* **9**, 458–463.

Hogarty, C.J., Ang, C. and Eitenmiller, R.R. (1989) Tocopherol content of selected foods by HPLC/fluorescence quantitation. *J. Food Compost. Anal.* **2**, 200–209.

James, C. *Global status of commercialized Biotech/GM crops*: 2012. *ISAAA Brief No. 44*. ISAAA: Ithaca, NY.

Privalle, L.S., Chen, J., Clapper, G., Hunst, P., Spiegelhalter, F. and Zhong, C.X. (2012) Development of an agricultural biotechnology crop product: testing from discovery to commercialization. *J. Agric. Food Chem.* **60**, 10179–10187.

Ridley, W.P., Sidhu, R.S., Pyla, P.D., Nemeth, M.A., Breeze, M.L. and Astwood, J.D. (2002) Comparison of the nutritional profile of glyphosate-tolerant corn event NK603 with that of conventional corn (*Zea mays* L.). *J. Agric. Food Chem.* **50**, 7235–7243.

Wehrhahn, C. and Allard, R.W. (1965) The detection and measurement of the effects of individual genes involved in the inheritance of a quantitative character in wheat. *Genetics*, **51**, 109–119.

Zhou, J., Berman, K.H., Breeze, M.L., Nemeth, M.A., Oliveira, W.S., Braga, D.P.V., Berger, G.U. and Harrigan, G.G. (2011a) Compositional variability in conventional and glyphosate-tolerant soybean (*Glycine max* L.) varieties grown in different regions in Brazil. *J. Agric. Food Chem.* **59**, 11652–11656.

Zhou, J., Harrigan, G.G., Berman, K.H., Webb, E.G., Klusmeyer, T.H. and Nemeth, M.A. (2011b) Stability in the composition equivalence of grain from insect-protected maize and seed from glyphosate-tolerant soybean to conventional counterparts over multiple seasons, locations, and breeding germplasms. *J. Agric. Food Chem.* **59**, 8822–8828.

Supporting information

Additional Supporting information may be found in the online version of this article:

Figure S1 Genetic distance and similarities of the conventional inbreds represented in this study.

Table S1 Starch protein and amino acid combined site mean values for grain from control hybrids.

Table S2 Starch protein and amino acid combined site mean values for grain from MON87460 hybrids.

Table S3 Starch protein and amino acid combined site mean values for grain from MON89034 hybrids.

Table S4 Starch protein and amino acid combined site mean values for grain from NK603 hybrids.

Table S5 Oil fatty acid and tocopherol combined site mean values for grain from control hybrids.

Table S6 Oil fatty acid and tocopherol combined site mean values for grain from MON87460 hybrids.

Table S7 Oil fatty acid and tocopherol combined site mean values for grain from MON89034 hybrids.
Table S8 Oil fatty acid and tocopherol combined site mean values for grain from NK603 hybrids.
Table S9 Mineral combined site mean values for grain from control hybrids.
Table S10 Mineral combined site mean values for grain from MON87460 hybrids.
Table S11 Mineral combined site mean values for grain from MON89034 hybrids.
Table S12 Mineral combined site mean values for grain from NK603 hybrids.
File S1 *Supporting_File1_Ind SiteData* provides individual site data in xml format.
File S2 *Supporting_File2_Histogram of Differences* provides graphical plots showing distribution of magnitude differences between near-isogenic comparators.
File S3 *Supporting_File3_Variance Component Plots* provides graphical plots showing the results of variance component analysis for all assessed crop compositional analytes.